

Dear Editor,

We thank you for sending our manuscript for review and giving us this opportunity to reply to the review comments. The comments were constructive and mostly asked for additional citations, clarifications and rephrasing of some text passages and improvements to the figures. We will provide an improved manuscript, addressing all review comments. Below are our detailed point-by-point replies and suggested manuscript improvements (blue) for each comment (black).

*In the name of all co-authors,
Markus Adloff*

Reviewer 1:

Markus Adloff and colleagues assess the sensitivity of the AMOC to changes in radiative forcing in the intermediate complexity model Bern3D. The range of the radiative forcing is representative of the last 800 kyr. The radiative forcing comprises orbital forcing, greenhouse gases, ice-sheet induced albedo changes and dust forcing. The strength of the radiative forcing is scaled by the maximum dust forcing at the LGM. The authors identify four stable AMOC states, a strong interglacial state, a weak glacial state and two less stable intermediate states. The magnitude of the radiative forcing determines the time the AMOC spends in each of the respective states. The authors analyse the characteristics of the AMOC states and assess the underlying mechanisms through further more idealised simulations. Comparison with available proxy data for sea surface temperature, AMOC strength and climate variability indicate that the simulations contain realistic AMOC behaviour (depending on the forcing strength) and that valuable insights on thermal AMOC thresholds throughout the glacial cycles can be obtained from them.

Overall the paper is of high quality, well written and definitely of interest for a wide audience in the CP community and beyond. Testing for thermal AMOC thresholds in itself is not new, but the length of the simulations and the large covered range of forcing scenarios that can only be achieved through the intermediate complexity of Bern3D provide enough novel insights.

My comments are mostly minor, asking for more clarification or context. I recommend publication after minor revisions.

We thank the reviewer for their time and effort, and the constructive comments, which helped to improve our manuscript.

Below are our detailed point-by-point replies and manuscript improvements (blue) for each comment (black). Line numbers refer to the new manuscript version without track changes.

Minor Comments:

Page 2:

I.3-4: This is the first time future AMOC stability is mentioned. It may be worthwhile to add a few sentences linking past and future AMOC stability.

We agree that the mention of future AMOC stability at this point is not well-connected to the rest of our manuscript. Since our study is only concerned with AMOC stability at pre-industrial and colder temperatures, we removed the sentences on future AMOC stability.

I.45: It would also be helpful to provide a bit more context on thermal thresholds. The previous two paragraphs mostly talk about the haline part (i.e. surface freshwater input and salinity redistribution). Which models have been used to analyse thermal AMOC thresholds and for which climate states? And could you comment on whether the AMOC in intermediate complexity models tends to be more or less or similarly stable as in fully coupled earth system models (e.g. the AMOC in ocean-only models is known to be more prone to instabilities than in coupled GCMs).

We added more information about the models used in studies investigating thermal forcings on AMOC, and provided more references (lines 95-109, 564-571). Further, we now explicitly mention that bistability of AMOC under thermal forcing has been observed in both coupled and uncoupled GCMs. The updated paragraph in the introduction will read:

Page 4:

I.16-17: Can you briefly explain why it is a useful approximation to use the LR04 stack as a scaling for the dust radiative forcing?

We now mention the close correlation of reconstructed dust fluxes with ice volume and provide a reference in the revised manuscript (lines 167-170).

Page6:

Fig1. Could you also show the combined radiative forcing of all three forcings? That would make it easier to identify periods of changing radiative forcing.

We added the combination of our dust forcing and the radiative forcing from greenhouse gases to Fig. 1. The orbital forcing does not cause substantial variations of the global radiation balance but rather the spatial and seasonal distribution of insolation. Hence, we preferred to keep showing the insolation changes at 65°N separately as an indication of how high-latitude radiative forcing evolved over the last glacial cycle.

Page7:

I.24-26: How do you assess stability here?

AMOC stability is a key concept for our study. However, we have not objectively defined it here, as the varying boundary conditions make it difficult to define an objective criterion that identifies the AMOC stability correctly in both full interglacial and glacial states. Instead, in the old version we chose to refer to 'stable modes', which are modes that are occupied by the AMOC most often, as diagnosed from Fig. 2. We agree that this is misleading and remove mentioning of stability from this paragraph, writing instead about the frequency of occurrence.

Page 8:

Fig.3 could you rotate the maps in the upper panels by 45°, so that the perspective on the North Atlantic becomes more easily comparable to the lower panels?

We changed the maps to focus on the North Atlantic region specifically.

I.22 and 26: Do you show stratification? The lower panels of Fig.3 only show surface density changes. Would it make sense to show stratification? Or do you infer increased stratification simply because of the lighter surface waters?

We changed the figure to include panels that show the density difference across the upper 1000 m of the water column as a metric for stratification.

Page 9:

I.5-10: I found this paragraph difficult to read and follow. If none of the differences is statistically significant, would it not be sufficient to report that MBT has no statistically significant effect on the AMOC response?

We shortened this paragraph by removing details about the differences, only mentioning their non-significance, as suggested by the reviewer (lines 316-321).

Page 11:

I.16-19: Does this refer to Fig.5 d? And in general: more specific references to Fig.5 could be made though out this page, to make it easier to follow. It is not always clear whether the text on this page refers to Fig.5, some other Figure or to results not shown.

We referenced figures more explicitly for the description of the processes.

I.20-21: Can you name the two processes and timescales explicitly? I guess they are N.Atl. freshwater changes (fast) and AABW propagation (slow), but it would be good to have them spelled out.

Yes, these were the processes we referred to. However, changes in the North Atlantic are more relevant for the observed AMOC changes, AABW propagation seems to have more of

a stabilising rather than destabilising effect. We removed this sentence and instead discuss changes in the North Atlantic and Southern Ocean separately.

Page 12:

I.20-24: I do not really see the further reduction in NADW export. To me, the distributions of all three water masses look almost identical at 23 and 24.5 kyr. Also, I do not really see NADW replacing AAIW, the upper NADW boundary does not seem to change and if anything, the southward extent of NADW also decreases.

We agree, our descriptions here were not accurate. We clarified this as follows (lines 439-442):

“The first abrupt shift in AMOC strength at 24.5 kyr in B.slow had only small effects on the water mass distribution. It mainly led to a reduced concentration of NADW at intermediate depths of the North Atlantic >45°N and a small increase of AABW concentration in the abyssal North Atlantic (Fig. 5d).”

Page 13:

I.25: What about the strong variance at 6kyr?

The strong variance at 6 kyr is associated with density changes in the North Atlantic. However, the AMOC appears to not undergo a state transition during this time. We added a description of this to the discussion of simulation B.slow to section 3.2 (lines 383-401).

Page 14:

I.32-40: Is this part meant in contrast to other models? The last sentence is also almost impossible to follow. Please consider a clearer formulation.

We rewrote this paragraph for a clearer discussion (lines 549-562).

Page 16/17:

Meta stable AMOC modes: How are the excitable/metastable states defined? By increased AMOC variance as in Fig.5? How do the metastable states relate to the four AMOC states I-IV from the beginning? Also, please consider adding the corresponding kyrs behind MIS3/4/5e etc, so that it is easier to identify the right parts of Fig.9 for those readers who do not have those numbers at the top of their heads.

We define excitable states as times when AMOC adopts intermediate modes II and III, which show more frequent AMOC strength shifts than the interglacial and glacial modes I and IV, respectively. We made this clearer in the text (lines 672-677). We also added the requested age information.

Page 18:

I.24-27: This would be very interesting indeed. I look forward to the follow-up :)

We too!

Technical/Editorial Comments:

Page 1

I.34: delete “boundary” after “Atlantic”

Deleted.

I.36-42: Very long and hard to read sentence. Consider reformulating for better readability. Also: which climate is being referred to at the end of the sentence? Probably North Pacific climate but it is not immediately clear.

We simplified and clarified this sentence as follows (lines 39-43):

“It also affects global climate by shifting the Intertropical Convergence Zone (ITCZ) and monsoon systems (Wang et al., 2001, Bozbiyik et al, 2011), and interacting with the regional climate and deep water formation in the North Pacific (Okazaki et al., 2010, Menviel et al., 2012, Praetorius and Mix, 2014).”

I.43-47: same as comment above. Also: Does the last half sentence (“and by modulating atmospheric greenhouse gas concentrations”) still correctly belong to the beginning of the sentence (“It influences deep ocean nutrient and oxygen concentrations”)?

We rewrote this sentence and clarified the role for greenhouse gas concentrations (lines 43-48). The new sentences read:

“The AMOC furthermore shapes biological surface productivity by regulating nutrient supply to the surface ocean in the Atlantic and Pacific (Tetard et al., 2017, Joos et al., 2017). On its southward path in the Atlantic, it influences deep ocean nutrient, carbon, and oxygen concentrations (Broecker, 1991). By affecting primary production and deep ocean carbon storage, AMOC changes also modulate atmospheric greenhouse gas concentrations (e.g. Menviel et al., 2008).”

Page 2:

I.2: “which had regional [...]” instead of “and had regional [...]”

Done.

Page 4:

I.13: What is meant with “rest of the past 800kyr”? Rest with respect to what? The spin up state?

This is a leftover from an amended sentence of a previous manuscript version. We removed “rest of the”.

Page 10:

Fig.5: Please increase the font size for better readability.

Done.

Page 12:

Fig.6: Please increase the font size for better readability.

Done.

Page 14:

I.23: The name of the ocean model is COCO (the ocean component of MIROC)

Done.

Page 15:

Fig.7: Please increase the font size for better readability.

Done.

Page 16:

I.28: wrong Figure reference? Should be Figure 1?

That is correct, we changed the figure reference accordingly to Fig. 1 in the revised manuscript.

Page 18:

I.14: delete “but”

Deleted.

Reviewer 2:

Summary

This study investigates the thermal thresholds of AMOC over the last 800kyr in the Bern3D climate model. By controlling the amplitude of the global cooling via dust forcing, they extract the timing of occurrences of the thermal threshold of AMOC. In addition, they manage to extract four stable and meta-stable AMOC modes, which differ in the locations of NADW formation regions and sea ice extent over the North Atlantic. The paper further explores mechanisms causing the AMOC weakening due to the reduction in the radiative forcing by means of hysteresis experiments. At the end of the paper, the authors provide a discussion on the limitations of the study by not including the orographic effect of the ice sheet in their simulations.

The experiments and results presented in this paper are very interesting (especially figs 2 and 9!) and are of interest of readers of the *Climate of the Past*. Therefore, I think this paper should be published. However, the explanation of the mechanism of the AMOC changes appears unclear to me (also noted by other reviewers). Additionally, there are some ambiguity in the experimental setup. Addressing these points would improve the overall quality of the manuscript. The comments are summarized below.

We thank the reviewer for the invested time in evaluating our study and the thoughtful comments that have helped to substantially improve the manuscript.

Below are our detailed point-by-point replies and suggested manuscript improvements (blue) for each comment (black). Line numbers refer to the new manuscript version without track changes.

Main comments:

Mechanisms by which the reduction of radiative forcing weakens the AMOC: Other reviewers have pointed out many points, so I'll just list potential ways to improve the manuscript.

1. Separate the paragraph explaining the effect of the Southern Ocean and North Atlantic. Section 3.2 goes back and forth between the role of NA and Southern Ocean. This makes it hard to follow the discussion. Related to this, Buizert and Schmittner (2015) provides a nice summary on the role of Southern Ocean. Ando and Oka (2021, GRL) also gives useful insight on the role of sea ice and heat transport on the stability of the AMOC. These two studies further performed hysteresis experiments with freshwater forcing. While the way of hysteresis experiment is not the same as in this study, I feel these studies should be cited and included in the discussion of the mechanism.

We rewrote the description of the processes at play in simulation B.slow (lines 370-419, 450-480), and followed the reviewer's advice to discuss changes in the North Atlantic and

Southern Ocean separately. We now also refer to Buizert and Schmittner (2015) and Ando and Oka (2021) in the discussion of our results (lines 564-599).

2. Use Fig. 3 to help explain the mechanism. For example, it would be more convincing for me if the authors explain the mechanism in the following manner “reduction of radiative forcing first weakens the convection in the Labrador Sea (Fig. 3) by increasing transport of sea ice from the arctic and by reducing the northward heat transport (Fig. 4). However, intensified surface cooling initiates the deepwater formation close to UK (Fig. 3), causing a shift of the AMOC into the second phase. Further reduction in radiative forcing” Obviously this is not a perfect example but please consider modifying the manuscript in this way.

We added more references to the figures, as suggested. We also clarified our description of processes (lines 366-504).

3. Relation of heat transport and the AMOC is always tricky. They vary together and also the heat transport can either weaken or strengthen the AMOC depending on the background condition (e.g. Paul and Schulz 2002, https://doi.org/10.1007/978-3-662-04965-5_5, Ando and Oka 2021, GRL). Please cite these paper when discussing the effect of heat transport on AMOC and explain why it should work in that sense.

We added to the discussion of processes changing AMOC strength in our simulations, and cited Ando and Oka, 2021 (lines 457-463):

“The simulated step changes in AMOC strength in our simulations were thus the response to gradual surface cooling and freshening, and occurred when NADW formation shifted southwards. The resulting redistributions of heat and salinity caused sudden shifts in the vertical density profiles and sea ice expansion which consolidated the new circulation mode (Ando and Oka, 2021). In particular, reduced advection of heat and salinity into former locations of deep water formation resulted in a more stable local water column (Fig. SI.7-9).”

Experimental setup: I think the authors need to explain why they decide to vary the magnitude of the dust related radiative forcing but not others in their sensitivity experiments (I'm not saying that's bad!). I don't fully understand how this model works, but isn't there another way to do similar experiments, e.g. changing the magnitude of the emissivity of the atmosphere or the magnitude of the ice sheet related radiative forcing? Effect of dust forcing is of course uncertain, but so are others (Tierney et al. 2020).

Related to 2, another question I have is that “Does the radiative forcing by dust affect the global and local temperatures in the same way as the GHG do in this model?” Looking at results from GCMs (e.g. Kawamura et al. 2017 Science Advances, Ohgaito et al. 2018 CP), it is shown that GHG and dust affect the local temperatures in a different way. This information is important especially when we want to use the insight from this study to better understand results of AOGCMs.

We clarified our methods and specifically noted that our applied forcing of radiation reductions are spatially uniform. As such, the pattern of the additional radiative forcing that

we prescribed is slightly different to that of GHG. GCM simulations showed that spatially different forcings lead to a very similar temperature pattern due to feedbacks (Boer, G. and Yu, B., 2003. Climate sensitivity and response. Climate Dynamics, 20, pp.415-429.). In either case, our simulations contain the radiative effect of GHG and the additional, uniform 'dust' forcing. Hence, we do not test the temperature effect of dust load changes specifically, but more generically of changes of the atmospheric radiation balance. There might be other ways of implementing this but in our model the effect would be virtually the same. We clarified our method description accordingly (lines 160-164, 182-186).

NADW formation in Norwegian/Greenland sea: This might be related to comments from other reviewers, but some previous studies have suggested the importance of cessation/resumption of convection over the Norwegian Sea when considering the thermal threshold of the AMOC (Oka et al. 2012). Please describe this feature in the Introduction and add some discussion wherever appropriate.

In the pre-industrial model state of Bern3D deep water formation does not occur north of the Irminger Sea. In the revised manuscript, we discuss the importance of the location of deep water formation sites at the beginning of the experiment for the existence of thermal thresholds. We added a discussion of this (lines 104-109, 564-571).

Specific comments:

P1L25-26: Given the limitations in the model, I think it would be safe to add "in this model" at the end of the sentence.

We specified that these results are only valid for our model in the revised manuscript.

P1L31-33: Isn't this the other way round; relatively salty water gets cooled by the atmosphere, the vertical density gradient weakens, and the water sinks and forms the NADW.

The reviewer is correct. We amended the sentence (lines 30-33):

"The Atlantic Meridional Overturning Circulation (AMOC) transports warm waters from the Southern Hemisphere and the Mexican Gulf towards the Nordic Seas, until the gradually cooled salty water lost enough buoyancy and sinks, forming North Atlantic Deep Water (NADW)."

P2L34: Perhaps "sensitive" -> "dependent"?

Done.

P3L13-25: So many references are missing in this paragraph. Please add the appropriate reference for each sentence. (e.g. references for Bern3D model, references for freshwater flux corrections)

We added the references describing the details of the Bern3D model and its setup as suggested.

P4L11-13: How did you define the maximum ice extent? Is it from the LGM?

We clarified our definition of the maximum forcing. Specifically, we changed our description in the Methods section (lines 165-167):

“The maximum radiative dust forcing, defined via the peak LGM value in the smoothed $\delta^{18}O$ stack, is a free parameter, ranging from 0 to -8 W/m^2 relative to PI (Simulations A.0 to A.8)”

P5L23-24: Better to say “stable”->“monostable”, “unstable”->“bistable” here.

With our stability tests, we assessed how resilient the circulation is to a small perturbation, i.e. whether it is close to a potential bifurcation point. However, we did not test each circulation state for mono- or bistability and the existence of bifurcation points. Hence, we do not think that the suggested terminology is appropriate at this point in the manuscript. Instead, we improved our terminology for our stability tests to clarify as follows (lines 482-485):

“Our stability experiments demonstrate that the circulation modes before and after the shifts recover from small freshwater perturbations, and can thus be considered as stable, i.e. sufficiently far from bifurcation points to recover from the small perturbation”

P11L11-15: I could not understand this sentence. Can you further elaborate on this, please?

We rewrote the description of simulation B.slow with a clearer discussion of the relevant processes. We described changes at the beginning of the simulation (lines 383-388):

“Initially, the whole Atlantic surface ocean cools and freshens, leaving the meridional temperature and salinity gradients almost unchanged (Fig 5e). However, NADW becomes less salty and colder in consequence (not shown) and the vertical density profiles in the subpolar North Atlantic change due to the temperature and salinity changes (Fig. SI.7-8).”

P13L4-5: Not quite sure what this positive feedback means here. In general, a surface cooling will reduce the SST and hence increase the surface density while the cooler SST reduces evaporation and causes a reduction in surface salinity and surface density. So isn't it a negative feedback?

We apologise for the confusion and the poorly formulated paragraph. We rewrote the paragraph for a clearer discussion of the relevant processes (lines 459-480). Instead of feedbacks, we wrote of stabilising processes, which is a clearer terminology.

P13L6: Isn't the sea ice feedback a positive feedback?

This is again an unclear description which we revised. In the new version, we mention sea ice expansion as a stabilising process (lines 528-536).

Figs.2 and 9: Very nice figures.

Thank you!

Reviewer 3:

In this manuscript, the authors performing several sets of transient experiments in Bern3D investigate thermally induced AMOC stability across glacial cycles. The results are new and complementary to our current theoretical understanding of glacial abrupt climate change. I believe this is a nice contribution to the community and suitable to Clim Past, but I reserve my recommendation for publication of this version since there remains room to improve its robustness and significance. In general, the authors shall 1) provide a more comprehensive introduction/discussion by considering at least most relevant literatures regarding AMOC stability during glacial cycles, 2) improve the clarity for mechanisms and feedback involved before, during and after AMOC transitions and 3) substantiate conclusions/statements by specifying the corresponding plots or adding direct modeling results/literatures. In addition, I would also recommend adding the 800-kyr results at least in the supplementary to provide an overview of the results, which would be of great interest for colleagues who are working on earlier glacial cycles as well.

We thank the reviewer for their time and effort reviewing our manuscript, and for the constructive comments which have helped to substantially improve our text. We addressed the reviewer's suggested improvements by completely rewriting our description and discussion of the processes that occur in the model, adding results to the figures in the main text and adding new figures to the SI, and referencing these figures more thoroughly in the text. Specifically, we made the changes outlined in our replies to the detailed comments below.

Below are our detailed point-by-point replies and suggested manuscript improvements (blue) for each comment (black). Line numbers refer to the new manuscript version without track changes.

Detailed comments are as follows:

P2L15-18: Freshwater input might be positive feedback to AMOC weakening as well. please refer to Barker et al 2015 and rephrase the sentences accordingly here as well as in L23-24.

We added the suggested reference and now mention the possibility for freshwater feedbacks. Specifically, we added the following sentence to our introduction (lines 72-74):

“Lags between the appearance of ice-rafted debris and the reconstructed cooling, however, suggest that freshwater fluxes could have instead acted as a positive feedback to AMOC weakening rather than triggering it (Barker et al., 2015).”

P2L25-29: other key relevant paper should be cited, for instance, Zhang et al., 2014, 2017.

We added references to Zhang et al., 2014 and 2017 and Vettoretti, 2022.

P2L33: also consider citing Zhang et al 2021, Vettoretti et al 2022 here.

Done.

P2L45-47: Please add relevant papers after the first sentence (e.g. Knorr and Lohmann 2007, Zhang et al 2017, Galbraith and de Lavergne 2018, etc.)

Done.

P4 L5: one predominant feature of glacial cycle is the development and demise of northern hemisphere ice sheet, involving both area and height, of which impacts on climate system are not the same. The former, as discussed in this study, via its albedo feedback is a thermal impact, while the latter, via its impacts on winds, is a kinetic impact (Zhang et al., 2014). In addition, there is no change in Bering Strait considered as well (Hu et al., 2011) (P5L4, a typo there). I was wondering how far these additional setups can alter the key messages of the thermal thresholds in this study. As seen in my following comments, at least a comprehensive discussion around this is required.

We added more discussion of other factors for AMOC stability (lines 584-599), see also our answers to further comments on this topic below.

P6 3.1: it would be good to present the 800kyr long transient simulation results. In Figure 1, it is of great help to add the radiative forcing curves to enable a comparison with B.slow experiment.

We added the according figure to the SI (Fig. SI.2).

P7L15-17: As alluded, lacking feedback from topo changes might overestimate the LGM cooling caused by radiative forcing decrease because higher NHIS can cause a stronger AMOC which promotes heat release from the ocean and hence North Atlantic warming. This might stimulate some discussion perhaps in data-model comparison or model limitation sections.

We added this to our discussion as suggested (lines 584-599).

P8. Fig3: given the North Atlantic and Nordic Sea are the key regions for AMOC state shift, it would be better to provide a zoom-in plot for this region, especially for the sea ice fraction

plot. Please also revise the color scheme for “sea ice cover fraction” to highlight change in the low values (<0.5) or just provide anomalous field as delta Density. Please also include lat-lon info in the plots. In addition, as you are discussing AMOC states, AMOC plots are highly recommended in this figure.

We changed the figure to show sea ice anomalies in the North Atlantic and added coordinates. We also added vector plots showing AMOC circulation to the SI. Further, we added AMOC plots to Fig 2.

P8L18-20: in the state (II), deep water formation is enhanced in west and south of Greenland. In general, it is more reddish in State (II) than in State (I), but why the AMOC is weakened in the former. Is this due to that convection in the western North Atlantic is not the key to the strength of the AMOC?

Yes, the mixed layer depth that we diagnosed from the annually-averaged model output is not a good metric to understand changes in AMOC strength. We explain the density changes more explicitly in the new manuscript version, and note that some changes in the locations of downwelling occur without changing AMOC strength (lines 379-395). We removed the plots of mixed layer depth and instead now show the absolute vertical density gradients in each state.

P8L25: “south-flowing fresh Arctic waters further stratify ...”. This is a key process to stabilize the glacial AMOC state, but in this version, there is not direct evidence to support it. Note that freshwater convergence in Fig5e cannot provide such support to this statement because it is a sum of freshwater flux across both 40N and 70N in the North Atlantic.

We now show the freshwater fluxes across each latitude separately in the updated Fig. 5. We also clarified that the spread of cold, fresh surface water in the North Atlantic stabilises the circulation state. We cannot actually determine whether the water comes from the Arctic or just turns more Arctic-like due to reduced influx of southern surface water. We rephrased the text accordingly.

P9L10: what is Kolmogorov-Smirnov test? Add details and reference.

We provided some more information and a reference (lines 319-321):

“none are statistically significant in the two-sided Smirnov test, which determines the likelihood that two distributions are the same (Berger and Zhou, 2014), even at the 50% confidence level”

P10 Fig 5: Panel e, it would be good to interpret meanings of positive/negative values of freshwater convergence to help readers understand this plot (e.g. positive values indicate freshwater import and hence a stable AMOC). In addition, the definition of freshwater convergence should be added to the Method section. It is worth noting that this AMOC stability indicator (Liu et al., 2014 Clim Dyn) predict a mono-stable AMOC regime in B.slow., in contrast to the hysteresis feature shown in Fig 4b. In addition, comparing the panel a) with

Fig 4b, it appears that B.slow.b is initialized from a AMOC state that is bistable with respect to radiative forcing. If so, why the AMOC recovers to its initial strong mode after removing the freshwater input? Typo in y-axis labels of panel c). it is also good to add radiative forcing panel on the top of it, with a vertical shaded bar to highlight periods when AMOC is bistable.

We amended Figure 5 as suggested. The hysteresis shown in Fig 4b is not the result of a traditional perturbation experiment with freshwater hosing but is the transient response to the applied radiative forcing. We are not sure if stability with regard to a freshwater perturbation is the same as stability in the face of changing boundary conditions, as caused by the radiative forcing. We are therefore careful with the interpretation of freshwater transport as stability indicators in our study. We added a discussion of freshwater transport to section 3.1 (lines 467-476):

“The difference between freshwater transport into the South Atlantic at 32°S and into the Arctic at 62.5°N in Fig. 5f can be used as a measure for the basin-wide salinity feedback (Rahmstorf, 1996, de Vries and Weber, 2005). In our simulation, changes in this metric were predominantly caused by changes in the transport across the northern edge, since transport into the South Atlantic remained almost unchanged throughout the cooling phase of B.slow. North Atlantic salinity is instead governed by changing transport from the subtropics into the North Atlantic and between the North Atlantic and Arctic. As such, the processes involved in the sudden AMOC strength changes, namely density changes in the upper water column, and those that stabilised new circulation modes (salinity and heat redistributions, sea ice expansion) mostly operated in the North Atlantic region.”

P11L31: how do you identify the reduced heat convergence “off the British Isles” based on the time series in Fig5?

The geographic information was not derived from the time series but provided as extra information to contextualise the time series. We added a figure to SI showing the discussed spatial pattern (new Fig. SI.11).

P11L33: It is also not logically clear why this is the cause to the northward spread of AABW. In Fig5, the northward intrusion of AABW is starting from the beginning of the experiment, not lagging the reduction of heat convergence in North Atlantic.

We agree, and changed the sentence to point out coincidence rather than causality. The experiment is initialised with no AABW tracer in the North Atlantic. Initially, changes in the concentration of the AABW tracer in the North Atlantic are small. Its amount only begins to rise substantially at ~15 kyr and shows the biggest jump at ~27 kyr when the heat convergence also declines.

P11L35: why “heat advection to >55N stops entirely”? could the authors present the evidence?

We apologise, this should have been ‘heat convergence’. We corrected this and added a figure of the spatial pattern of heat convergence changes to the SI (new Fig. SI.11).

P11L37-39: Again, no direct lines of evidence to support this statement. Does the contemporary sea ice expansion and its seasonality contribute to the freshening in the eastern Nordic Sea? As well as in P11L42-43. Please clarify. Is there a bipolar thermal seesaw during abrupt AMOC reduction in B.slow? The results appear to show that bipolar sea ice change out of phase with AMOC/NADW change – sea ice expansion with NADW weakening. The subdued thermal seesaw in B.slow indicates the dominant role of decreasing radiative forcing in controlling bipolar change.

Yes, we thank the reviewer for pointing this out. There is a small bipolar seesaw effect. We mention this explicitly in the revised manuscript. We also added a plot of with spatial patterns of changes in B.slow to the SI and add the following text to section 3.1 (lines 410-415):

“The biggest AMOC weakening at ~27 kyr is also accompanied by a weak bipolar seesaw effect, which causes a temporary decline in sea ice coverage in the Atlantic sector of the Southern Ocean (Fig. 5). It is, however, too small to reduce the radiation-driven sea ice increase in the longer term.”

P12L27: what’s the statement “... increased heat advection into the North Atlantic” based on?

This statement was erroneous and we deleted it. AMOC strength is not constant at the beginning of the experiment but weakens slowly, while the spatial pattern of deep convection and heat convergence in the North Atlantic change. We rewrote our description of the processes responsible for AMOC changes and use more references to figures. We added the spatial patterns of heat convergence changes to the new Fig. SI.11.

P12L29: weakened north ward transport of what? Upper cell of the AMOC?

We clarified that we mean the transport of salt and heat.

P13L1: please show the weakened the meridional salinity gradient in the North Atlantic.

We apologise, this was meant to say weakened meridional salinity transport, i.e. an increased salinity gradient. We corrected the statement in the text and added the temporal evolution of salinity in the Irminger and Caribbean Seas to Fig. 5 to show the increased meridional salinity gradient.

P13L3-5: how does the increased surface density promote SST decrease? This is not clear at all here.

We agree with the reviewer that this formulation was misleading. We meant to express that SST changes have a direct effect on water density and an indirect one via influencing evaporation, and that temperature-driven evaporation changes counteract the buoyancy forcing caused by the temperature change. We rewrote this paragraph as follows (lines 459-466):

“The deep water formation regions are sensitive to heat and salt flux changes, because any reduction in sea surface temperatures (SST) increases surface density but simultaneously reduces evaporation in ice-free areas, thus effectively creating a small freshwater forcing and a negative feedback to the buoyancy changes caused by the initial SST decrease. Sea ice covering the downwelling areas stabilises the water column by preventing surface ocean cooling and evaporation. The progressive influx of AABW into the North Atlantic is a further process stabilising new circulation modes by stratifying the water column from below (Buizert and Schmittner, 2015).”

P13L5-7: the authors proposed that sea ice expansion over convection sites acts as negative feedback in response to SST cooling, which is not convincing. This process, as demonstrated in this sentence, can avoid further cooling of sea surface, which in turn reduced sea surface heat loss to increase surface density, and thus stratifying the water column. This seems to exert rather positive feedback to stabilizing the cooling-induced AMOC slowdown. Please clarify. In general, positive/negative feedback discussed in this paragraph is hard to follow. Please clarify with more direct evidence/references.

We agree with the reviewer that the discussion of feedbacks in this paragraph was unclear. Importantly, the old version suggested that sea ice cover is a negative feedback on SST changes, which is not correct. Sea ice cover prevents evaporation and heat loss to the atmosphere, stabilising the water column. We rewrote the paragraph focussed on stabilising mechanisms rather than feedbacks (lines 455-476).

P13L11-13: As mentioned in previous comments, providing supportive evidence is of crucial importance since this is important positive feedback to the AMOC slow-down.

We made more references to figures in the re-written paragraph.

P13L15: please clarify and specify the positive and negative feedback mentioned here.

We agree that this paragraph was unclear, and we rewrote it by using the clearer term ‘stabilising process’ (see above).

P13L22-23: given the gradual decreasing radiative forcing, it is not clear whether it is the self oscillation or just an increased variability (small magnitude, 0.5Sv) as the system approaches the threshold. It appears that AMOC variance is of comparable or even larger magnitude during 6- 11kyr (Fig 5b). Is this also corresponding to self-oscillation?

We are now more cautious with our statement and only write that variability is increased. The large variability at 6-11 kyr is related to density changes in the Irminger Sea. We discuss this in the new manuscript version in lines 386-393 in section 3.1, but we didn’t see indications of oscillations:

“After about 6 kyr, the changes in the North Atlantic density profile shifted the location of NADW formation: NADW formation moved south as vertical density profiles in the subpolar

east North Atlantic stabilised under a freshening of the surface and density profiles further south steepened due to surface cooling combined with subsurface warming (Fig. SI.7-9). These changes did not cause a step-change in AMOC strength, but freshwater and heat advection into the North Atlantic was reduced, sea ice expansion increased in the eastern North Atlantic, and AMOC variance (calculated over a moving 50-year window) was increased (Fig. 5).”

P13P26-33: as discussed, results from B.slow.b seem not to support the hysteresis behavior with respect to radiative forcing change. What about stability/sensitivity of the AMOC at ~6kyr in B.slow?

We referred to hysteresis behaviour here because the radiative forcing that is required to cause the abrupt weakening of AMOC is not the same as the forcing required for strengthening it again. We are clearer in the revised manuscript. We did not test AMOC stability explicitly at 6 kyr because the high variability seems to cease once the density field has re-adjusted.

P13L36: Orbital configuration consists of three orbital parameters. Their combinations in the chosen time slices are different but this does not mean the associated climatic impacts are significantly distinct, for instance, 21ka versus 0ka. It is thus better to show values of obliquity, precession, eccentricity and boreal summer insolation for the chosen time slices here, which would be helpful to clarify whether orbital forcing matters the transient behavior of the AMOC. A better approach to test roles of orbital configurations is to re-conduct such transient experiments based on orbital sensitivity, for example, high versus low obliquity experiments (e.g. experiments in Extended Data Table 1 of Zhang et al 2021).

We now provide the orbital parameter values for each experiment in the SI. We were mostly interested here to see that changing the orbital configuration does not substantially alter the simulation results. It would be interesting to investigate the role of orbital changes for thermal thresholds in more detail in the future.

P14L5-7: not a full list of key relevant papers. Please add Knorr & Lohmann 2007, Banderas et al 2012 and Zhang et al 2017. Re multiple stable AMOC states, the difference in the strength of the AMOC is significantly different with a magnitude of >5Sv. In this context, it appears that the metastable AMOC states proposed here are perhaps sub-states of the interglacial/glacial AMOC state. Given the low AMOC variability in Bern3D, I assume this might not be reproducible by full GCMs nor perhaps in proxies.

We thank the reviewer for these additional relevant references, which we added to the paragraph. We also added further SI figures that show that each of the four persistent AMOC strengths is associated with different Greenland temperatures and North Atlantic sea ice extents, suggesting that they correspond to different climate states. Further, we agree, that it would be interesting to see this tested with a full GCM in future studies

P14L12-16: what's the exact role of 'heat advection" in AMOC mode transition? A positive feedback, a trigger or else? It would be good to have a clearer description here to specify the importance of heat advection.

Yes, we clarified this. Changes in heat convergence only seem to stabilise density profiles in the North Atlantic rather than causing AMOC shifts. We added these two paragraphs to section 3.1 (lines 386-397 and 453-457):

“After about 6 kyr, the changes in the North Atlantic density profile shifted the location of NADW formation: NADW formation moved south as vertical density profiles in the subpolar east North Atlantic stabilised under a freshening of the surface and density profiles further south steepened due to surface cooling combined with subsurface warming (Fig. SI.7-9). These changes did not cause a step-change in AMOC strength, but freshwater and heat advection into the North Atlantic was reduced, sea ice expansion increased in the eastern North Atlantic, and AMOC variance (calculated over a moving 50-year window) was increased (Fig. 5). Transport of heat and salinity into the North Atlantic decreased (Fig. 5f, g), which reduced North Atlantic SST and SSS (Fig. 5e). The reduced influx of subtropical surface waters also caused abrupt cooling and freshening in the Irminger Sea (Fig. SI.8). At 24 kyr, the AMOC had weakened to ~14.5 Sv and sea ice cover extended south of the Irminger Sea (Fig SI.11).”

“The simulated step changes in AMOC strength in our simulations were thus the response to gradual surface cooling and freshening, and occurred when NADW formation shifted southwards. The resulting redistributions of heat and salinity caused sudden shifts in the vertical density profiles and sea ice expansion which consolidated the new circulation mode (Ando and Oka, 2021).”

P14L26-27: it is not true. For instance, Zhang et al 2017 applying a fully coupled AOGCM proposes that atmospheric CO₂ levels are of control for glacial AMOC bi-stability.

Our intention here was to understand why Oka et al. (2021), specifically, required a stronger forcing in the southern hemisphere for thermal thresholds to arise, while we see thermal thresholds in our model under a globally uniform forcing. We clarified this and now also refer to Zhang et al. (2017) (lines 534-595).

P14L28-30: this may be true if comparing with other EMICs or simple models but not for GCM. Please clarify.

This was a wrong conception, we removed this statement from the manuscript.

P14L35: please provide modeling results or relevant literatures to support this statement especially regarding poleward moisture transport. It appears to me Fig 5e would be the right panel to refer to given the different trends between Atlantic and North Atlantic freshwater convergence. Sentences in P11L13-14 seem already touch this point, but it requires future clarification to link them to moisture transport and so on.

The statement on moisture transport was specific to our model. The wind field is constant but the water holding capacity of air decreases with the temperature decline. Hence, less moisture is transported polewards by the large-scale atmospheric circulation. We clarified this in the text (lines 543-558). We also added more specific metrics of the changing water balance to the figures. We added SSS and SST timeseries for the Caribbean and Irminger Sea, as well as marine freshwater fluxes across latitudes 37.5°N and 62.5°N to Fig. 5, and spatial changes of P-E in the North Atlantic to the SI.

P15: it is good to see the discussion about potential impacts of other parameters, especially ice sheet topography and associated wind, on the simulated AMOC change in different transient runs. In a glacial cycle, both changes in radiative forcing (e.g. CO₂) and wind circulation/gateway caused by ice volume changes play a role in the strength/stability of the AMOC (Hu et al., 2011, Zhang et al 2014, 2017, 2021). Of most relevance here is their opposite impacts on the strength of the AMOC through glacial cycles in comparison to the thermal forcing (Barker and Knorr 2021). In this study, the authors investigated the roles of changes in radiative forcing in AMOC stability, which is the half story of AMOC multi-equilibria in glacial cycles. How do changes in those key parameters influence the results of A experiments? I would be happy to see more comprehensive discussion around this here as well as in Section 3.4 and 3.5. Perhaps, Section 3.3-3.5 can be integrated to one section to highlight and discuss the current understanding of AMOC stability, impacts of current model limitation on the current results and data-model comparison, and their implications and future perspectives.

We followed the reviewer's advice and added more discussion of AMOC stability and model limitations by combining sections 3.3 and 3.5 (lines 509-664).

P15 L14-15: Please add relevant reference to “different representations of processes affecting AABW density changes”.

We added a reference here (line 627): “(e.g. brine rejection, Bouttes et al., 2011)”

P17 Figure 9: it would be good to flip y-axis of d18O curve upside down, given the tradition of plotting LR04/sea level curves.

We inverted the y-axis of the $\delta^{18}\text{O}$ panel as suggested.

References

Ando, T. and Oka, A., 2021. Hysteresis of the glacial Atlantic meridional overturning circulation controlled by thermal feedbacks. *Geophysical Research Letters*, 48(24), p.e2021GL095809.

Arzel, O., England, M.H. and Sijp, W.P., 2008. Reduced stability of the Atlantic meridional overturning circulation due to wind stress feedback during glacial times. *Journal of climate*, 21(23), pp.6260-6282.

Banderas, R., Álvarez-Solas, J. and Montoya, M., 2012. Role of CO₂ and Southern Ocean winds in

glacial abrupt climate change. *Climate of the Past*, 8(3), pp.1011-1021.

Barker, S., Chen, J., Gong, X., Jonkers, L., Knorr, G. and Thornalley, D., 2015. Icebergs not the trigger for North Atlantic cold events. *Nature*, 520(7547), pp.333-336.

Berger, V.W. and Zhou, Y., 2014. Kolmogorov–smirnov test: Overview. *Wiley statsref: Statistics reference online*.

Bouttes, N., Paillard, D., Roche, D.M., Brovkin, V. and Bopp, L., 2011. Last Glacial Maximum CO₂ and $\delta^{13}\text{C}$ successfully reconciled. *Geophysical Research Letters*, 38(2).

Bozbiyik, A., Steinacher, M., Joos, F., Stocker, T.F. and Menviel, L., 2011. Fingerprints of changes in the terrestrial carbon cycle in response to large reorganizations in ocean circulation. *Climate of the Past*, 7(1), pp.319-338.

Broecker, W.S., Blanton, S., Smethie Jr, W.M. and Ostlund, G., 1991. Radiocarbon decay and oxygen utilization in the deep Atlantic Ocean. *Global Biogeochemical Cycles*, 5(1), pp.87-117.

Buizert, C. and Schmittner, A., 2015. Southern Ocean control of glacial AMOC stability and Dansgaard-Oeschger interstadial duration. *Paleoceanography*, 30(12), pp.1595-1612.

de Vries, P. and Weber, S.L., 2005. The Atlantic freshwater budget as a diagnostic for the existence of a stable shut down of the meridional overturning circulation. *Geophysical Research Letters*, 32(9).

Galbraith, E. and de Lavergne, C., 2019. Response of a comprehensive climate model to a broad range of external forcings: relevance for deep ocean ventilation and the development of late Cenozoic ice ages. *Climate Dynamics*, 52, pp.653-679.

Ganopolski, A. and Rahmstorf, S., 2001. Rapid changes of glacial climate simulated in a coupled climate model. *Nature*, 409(6817), pp.153-158.

Gregory, J.M., Dixon, K.W., Stouffer, R.J., Weaver, A.J., Driesschaert, E., Eby, M., Fichefet, T., Hasumi, H., Hu, A., Jungclaus, J.H. and Kamenkovich, I.V., 2005. A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO₂ concentration. *Geophysical Research Letters*, 32(12).

Hu, A., Meehl, G.A., Han, W., Timmermann, A., Otto-Bliesner, B., Liu, Z., Washington, W.M., Large, W., Abe-Ouchi, A., Kimoto, M. and Lambeck, K., 2012. Role of the Bering Strait on the hysteresis of the ocean conveyor belt circulation and glacial climate stability. *Proceedings of the National Academy of Sciences*, 109(17), pp.6417-6422.

Jackson, L.C. and Wood, R.A., 2018. Hysteresis and resilience of the AMOC in an eddy-permitting GCM. *Geophysical Research Letters*, 45(16), pp.8547-8556.

Jackson, L.C., Alastrué de Asenjo, E., Bellomo, K., Danabasoglu, G., Haak, H., Hu, A., Jungclaus, J.H., Lee, W., Meccia, V.L., Saenko, O. and Shao, A., 2023. Understanding AMOC stability: the North Atlantic hosing model intercomparison project. *Geoscientific Model Development*, 16, pp.1975-1995.

Joos, H., Madonna, E., Witlox, K., Ferrachat, S., Wernli, H. and Lohmann, U., 2017. Effect of anthropogenic aerosol emissions on precipitation in warm conveyor belts in the western North Pacific in winter—a model study with ECHAM6-HAM. *Atmospheric chemistry and physics*, 17(10), pp.6243-6255.

Klockmann, M., Mikolajewicz, U. and Marotzke, J., 2018. Two AMOC states in response to decreasing greenhouse gas concentrations in the coupled climate model MPI-ESM. *Journal of Climate*, 31(19), pp.7969-7984.

Knorr, G. and Lohmann, G., 2007. Rapid transitions in the Atlantic thermohaline circulation triggered

- by global warming and meltwater during the last deglaciation. *Geochemistry, Geophysics, Geosystems*, 8(12).
- Li, C. and Born, A., 2019. Coupled atmosphere-ice-ocean dynamics in Dansgaard-Oeschger events. *Quaternary Science Reviews*, 203, pp.1-20.
- Lohmann, J., Dijkstra, H.A., Jochum, M., Lucarini, V. and Ditlevsen, P.D., 2023. Multistability and Intermediate Tipping of the Atlantic Ocean Circulation. arXiv preprint arXiv:2304.05664.
- Menary, M.B., Roberts, C.D., Palmer, M.D., Halloran, P.R., Jackson, L., Wood, R.A., Müller, W.A., Matei, D. and Lee, S.K., 2013. Mechanisms of aerosol-forced AMOC variability in a state of the art climate model. *Journal of Geophysical Research: Oceans*, 118(4), pp.2087-2096.
- Menviel, L., Timmermann, A., Mouchet, A. and Timm, O., 2008. Meridional reorganizations of marine and terrestrial productivity during Heinrich events. *Paleoceanography*, 23(1).
- Menviel, L., Joos, F. and Ritz, S.P., 2012. Simulating atmospheric CO₂, 13C and the marine carbon cycle during the Last Glacial–Interglacial cycle: possible role for a deepening of the mean remineralization depth and an increase in the oceanic nutrient inventory. *Quaternary Science Reviews*, 56, pp.46-68.
- Mikolajewicz, U., Santer, B.D. and Maier-Reimer, E., 1990. Ocean response to greenhouse warming. *Nature*, 345(6276), pp.589-593.
- Oka, A., Hasumi, H. and Abe-Ouchi, A., 2012. The thermal threshold of the Atlantic meridional overturning circulation and its control by wind stress forcing during glacial climate. *Geophysical Research Letters*, 39(9).
- Oka, A., Abe-Ouchi, A., Sherriff-Tadano, S., Yokoyama, Y., Kawamura, K. and Hasumi, H., 2021. Glacial mode shift of the Atlantic meridional overturning circulation by warming over the Southern Ocean. *Communications Earth & Environment*, 2(1), p.169.
- Okazaki, Y., Timmermann, A., Menviel, L., Harada, N., Abe-Ouchi, A., Chikamoto, M.O., Mouchet, A. and Asahi, H., 2010. Deepwater formation in the North Pacific during the last glacial termination. *Science*, 329(5988), pp.200-204.
- Pöppelmeier, F., Scheen, J., Jeltsch-Thömmes, A. and Stocker, T.F., 2021. Simulated stability of the Atlantic meridional overturning circulation during the Last Glacial Maximum. *Climate of the Past*, 17(2), pp.615-632.
- Praetorius, S.K. and Mix, A.C., 2014. Synchronization of North Pacific and Greenland climates preceded abrupt deglacial warming. *Science*, 345(6195), pp.444-448.
- Rahmstorf, S., 1996. On the freshwater forcing and transport of the Atlantic thermohaline circulation, *Clim. Dyn.*, 12, 799–811.
- Sherriff-Tadano, S., Abe-Ouchi, A., Yoshimori, M., Ohgaito, R., Vadsaria, T., Chan, W.L., Hotta, H., Kikuchi, M., Kodama, T., Oka, A. and Suzuki, K., 2023. Southern Ocean surface temperatures and cloud biases in climate models connected to the representation of glacial deep ocean circulation. *Journal of Climate*, 36(11), pp.3849-3866.
- Stommel, H., 1961. Thermohaline convection with two stable regimes of flow. *Tellus*, 13(2), 224– 230. <https://doi.org/10.3402/tellusb.v13i2.12985>

- Tetard, M., Licari, L. and Beaufort, L., 2017. Oxygen history off Baja California over the last 80 kyr: A new foraminiferal-based record. *Paleoceanography*, 32(3), pp.246-264.
- Vettoretti, G., Ditlevsen, P., Jochum, M. and Rasmussen, S.O., 2022. Atmospheric CO₂ control of spontaneous millennial-scale ice age climate oscillations. *Nature Geoscience*, 15(4), pp.300-306.
- Wang, Y.J., Cheng, H., Edwards, R.L., An, Z.S., Wu, J.Y., Shen, C.C. and Dorale, J.A., 2001. A high-resolution absolute-dated late Pleistocene monsoon record from Hulu Cave, China. *Science*, 294(5550), pp.2345-2348.
- Weijer, W., Cheng, W., Drijfhout, S.S., Fedorov, A.V., Hu, A., Jackson, L.C., Liu, W., McDonagh, E.L., Mecking, J.V. and Zhang, J., 2019. Stability of the Atlantic Meridional Overturning Circulation: A review and synthesis. *Journal of Geophysical Research: Oceans*, 124(8), pp.5336-5375.
- Weijer, W., Cheng, W., Garuba, O.A., Hu, A. and Nadiga, B.T., 2020. CMIP6 models predict significant 21st century decline of the Atlantic meridional overturning circulation. *Geophysical Research Letters*, 47(12), p.e2019GL086075.
- Winckler, G., Anderson, R.F., Fleisher, M.Q., McGee, D. and Mahowald, N., 2008. Covariant glacial-interglacial dust fluxes in the equatorial Pacific and Antarctica. *science*, 320(5872), pp.93-96.
- Yang, H., Wang, K., Dai, H., Wang, Y. and Li, Q., 2016. Wind effect on the Atlantic meridional overturning circulation via sea ice and vertical diffusion. *Climate Dynamics*, 46, pp.3387-3403.
- Zhang, X., Prange, M., Merkel, U. and Schulz, M., 2014. Instability of the Atlantic overturning circulation during Marine Isotope Stage 3. *Geophysical Research Letters*, 41(12), pp.4285-4293.
- Zhang, X., Knorr, G., Lohmann, G. and Barker, S., 2017. Abrupt North Atlantic circulation changes in response to gradual CO₂ forcing in a glacial climate state. *Nature Geoscience*, 10(7), pp.518-523.